

utterances: 506

tokens total: 12,881

id/en tokens: 12,510 | unk: 371

p(ID)=0.251, p(EN)=0.749

M-index=0.604

Language entropy=0.813 bit

CMI (corpus)=25.14

I-index mean=0.383

langid_threshold=0.80



A stacked horizontal bar chart showing the distribution of tokens across three categories: ID (blue), EN (orange), and UNK (green). The total length of the bar is approximately 12,881 units. The ID category accounts for about 3,100 units, EN for about 9,500 units, and UNK for about 200 units.

ID EN UNK

12000

10000

8000

6000

4000

2000

0

tokens